

Preparing code and data for computational reproducibility: a hands-on workshop

Computational analyses are playing an increasingly central role in research. Journals, funders, and researchers are calling for published research to include associated data and code.

However, many involved in research have not received training in best practices and tools for sharing code and data. This workshop aims to address this gap in training while also providing those who support researchers with curated best practices guidance and tools.

This workshop is unique compared to other reproducibility workshops due to its practical, step-by-step design. It is comprised of hands-on exercises to prepare research code and data for computationally reproducible publication. Although the workshop starts with some brief introductory information about computational reproducibility, the bulk of the workshop is guided work with data and code. The basic best practices for publishing code and data are covered with curated resources and participants move through preparing research for reuse, organization, documentation, automation, and submitting their code and data to share. Tools to support reproducibility will be introduced but all lessons will be platform agnostic.

Workshop goals

- Learn best practices for file organization, documentation, automation, and dissemination.
- Assess possible tools for publishing code and data.
- Submit your code and data for sharing.

Best practices topics covered: Organization

- Create one repository or directory that holds all related research files.
- Organize your research to separate data, code, and results.
- Save results explicitly.

Documentation

- Document each element or variable in your dataset with a codebook.
- Create a project README file.
- Specify licenses for your data and your code.
- Use literate programming.
- Specify your computational environment and package versions.

Automation

- Use container technology to make your analysis portable and reusable.
- Change absolute paths to relative paths.
- Create a master script for your analyses.

Dissemination

- Write a detailed study protocol before you gather your data.
- Report all results, no matter their direction or statistical significance.
- Publish and share your data and code.



22nd International Symposium on Electronic Theses and Dissertations

Preparation

Participants should bring a laptop to fully participate. Participants may bring their own data and code to work through during the workshop. If they do not have code and data of their own to bring, they will follow along with example code and data.

Audience

The audiences for this course are researchers and research support staff who are involved in the preparation and publication of research materials. Anyone with an interest in reproducible publications is welcome. This course is especially useful for those looking to learn practical steps in improving the computational reproducibility of their own research.